Fact Table: **FeedbackFact**

Granularity: Transaction

Grain Definition: Each row consists of one instance of feedback for one student/class given by one teacher based on one observed skill/behavior

Associated Dimensions: The suitable dimensions are DATE, STUDENT, SKILL, TEACHER, SCHOOL, and CLASS.

Fact Table: **AttendanceFact**

Granularity: Transaction

Grain Definition: Each row consists of one attendance status taken for one student in a given class at a given time

Associated Dimensions: The suitable dimensions are DATE, STUDENT, SCHOOL, and CLASS.

--- ISQA8410 Project Part 3, SQL queries

--- Eric Troudt

--- 1. List the number of feedbacks broken down by class and student for the school year 2020-2021 for one

--- school. The output should list {classroom, student, number of feedback}.

SELECT cd\_cl.class\_name classroom, cd\_st.student\_key student, COUNT(cd\_fd.feedback\_id) "number of feedback"

FROM ClassDojo\_FeedbackFact cd\_fd

JOIN ClassDojo\_ClassDim cd\_cl ON cd\_fd.class\_key = cd\_cl.class\_key

JOIN ClassDojo\_StudentDim cd\_st ON cd\_fd.student\_key = cd\_st.student\_key

JOIN ClassDojo\_DateDim cd\_dt ON (cd\_fd.date\_key = cd\_dt.date\_key AND cd\_dt.date\_academic\_year = '2020-2021')

JOIN ClassDojo\_SchoolDim cd\_sch ON (cd.fd.school\_key = cd\_sch.school\_key AND cd\_sch.school\_name = '\*\*a given school name\*\*')

GROUP BY cd\_cl.class\_name, cd\_st.student\_key;

--- 2. List the number of positive and negative feedback broken down by teacher of one school in school year

--- 2020-2021. The output should list {teacher, number of positive feedback, number of negative feedback}.

SELECT cd\_t.teacher\_id, COUNT(CASE WHEN cd\_sk.negpos = 'pos' THEN 1 END) "# of positive feedback",

COUNT(CASE WHEN cd\_sk.negpos = 'neg' THEN 1 END) "# of negative feedback"

FROM ClassDojo\_FeedbackFact cd\_fd

JOIN ClassDojo\_TeacherDim cd\_t ON cd\_fd.teacher\_key = cd\_t.teacher\_key

JOIN ClassDojo\_SkillDim cd\_sk ON cd.fd.skill\_id = cd\_sk.skill\_id

JOIN ClassDojo\_DateDim cd\_dt ON (cd\_fd.date\_key = cd\_dt.date\_key AND cd\_dt.date\_academic\_year = '2020-2021')

JOIN ClassDojo\_SchoolDim cd\_sch ON (cd.fd.school\_key = cd\_sch.school\_key AND cd\_sch.school\_name = '\*\*a given school name\*\*')

GROUP BY cd\_t.teacher\_id;

--- 3. List the feedback points for one student for all the months in 2020. The output should list {month, class,

--- feedback points}

SELECT cd\_dt.date\_month "month", cd\_cl.class\_name "class", SUM(cd\_fd.feedback\_points) "Feedback Total Points"

FROM ClassDojo\_FeedbackFact cd\_fd

JOIN ClassDojo\_DateDim cd\_dt ON cd\_fd.date\_key = cd\_dt.date\_key

JOIN ClassDojo\_ClassDim cd\_cl ON cd\_fd.class\_key = cd\_cl.class\_key

JOIN ClassDojo\_StudentDim cd\_st ON (cd\_fd.student\_key = cd\_st.student\_key AND cd.st.student\_key = "\*\* given student key\*\*")

GROUP BY cd\_dt.date\_month, cd\_cl.class\_name;

--- 4. List the number of feedbacks broken down by skill for one class for each academic quarter in academic

--- year 2020-2021. The output should list {skill, quarter, number of feedbacks}

SELECT cd\_sk.skill\_name "skill name", cd\_dt.date\_academic\_year\_quarter "academic quarter (2020-2021)", COUNT(cd\_fd.feedback\_id)

FROM ClassDojo\_FeedbackFact cd\_fd

JOIN ClassDojo\_SkillDim cd\_sk ON cd.fd.skill\_id = cd\_sk.skill\_id

JOIN ClassDojo\_DateDim cd\_dt ON (cd\_fd.date\_key = cd\_dt.date\_key AND cd\_dt.date\_academic\_year = '2020-2021')

JOIN ClassDojo\_ClassDim cd\_cl ON cd\_fd.class\_key = cd\_cl.class\_key AND cd\_cl.class\_name = '\*\*a given class\*\*'

GROUP BY cd\_sk.skill\_name, cd\_dt.date\_academic\_year\_quarter;

--- 5. List the monthly attendance status for students in a class for school year 2020-2021. The output should list

--- {student, month, status, number of attendance}

SELECT cd\_st.student\_key student, cd\_dt.date\_month "month", cd\_at.attendance\_status, COUNT(attendance\_status) "# of attendance"

FROM ClassDojo\_AttendanceFact cd\_at

JOIN ClassDojo\_StudentDim cd\_st ON cd\_at.student\_key = cd\_st.student\_key

JOIN ClassDojo\_DateDim cd\_dt ON (cd\_at.date\_key = cd\_dt.date\_key AND cd\_dt.date\_academic\_year = '2020-2021')

JOIN ClassDojo\_ClassDim cd\_cl ON cd\_at.class\_key = cd\_cl.class\_key AND cd\_cl.class\_name = '\*\*a given class\*\*'

GROUP BY cd\_st.student\_key, cd\_dt.date\_month, cd\_at.attendance\_status;

--- 6. List the attendance status for students per school districts for schools in year 2020-2021. The output

--- should list {district, status, number of attendance}

SELECT cd\_sch.school\_district\_name, cd\_at.attendance\_status, COUNT(attendance\_status) "# of attendance"

FROM ClassDojo\_AttendanceFact cd\_at

JOIN ClassDojo\_SchoolDim cd\_sch ON cd.at.school\_key = cd\_sch.school\_key

JOIN ClassDojo\_DateDim cd\_dt ON (cd\_at.date\_key = cd\_dt.date\_key AND cd\_dt.date\_academic\_year = '2020-2021')

JOIN ClassDojo\_ClassDim cd\_cl ON cd\_at.class\_key = cd\_cl.class\_key

GROUP BY cd\_sch.school\_district\_name, cd\_at.attendance\_status;

What have I got out of doing this assignment? How have I developed my knowledge and skills? How do I see the payoff from doing this assignment helping me in the longer term?

For me, this assignment (and also week 12, 13) gave me a nice overview of how a data warehouse is built from source data in OLTP systems. By working through the process of identifying a fact table along with its corresponding dimensions from which sql queries could be written, I was able to gain a stronger appreciation for the necessity of data marts/warehouses for supporting the analytical needs of business intelligence. I think it’s important to have a comprehensive introduction to warehouses because they are the foundations for businesses needing to extract vital information from disparate sources of data.

To what extent has this assignment helped me to clarify what I need to learn about this topic? Have I a clearer picture after doing the assignment, or a foggier one? Who can help me gain a clearer picture, if the latter?

I’ve always been really interested in the various applications and tools being developed for big data environments, but it’s difficult to understand what purposes they are serving without at least having a general understanding of the infrastructure (data warehouses, data lakes, etc.) that are developed to accommodate data on such a large scale in terms of both volume and heterogeneity.